SUPPLEMENTAL MATERIAL FOR 'GENETIC EFFECTS ON TOXIC AND ESSENTIAL ELEMENTS IN HUMANS: ARSENIC, CADMIUM, COPPER, LEAD, MERCURY, SELENIUM AND ZINC IN ERYTHROCYTES'

MATERIALS AND METHODS

Participants. All participants were twins enrolled in the Australian Twin Registry, born between 1903 and 1964. The subjects and procedures were the same as those described for blood lead in a previous publication (Whitfield et al., 2007). Participants had completed a postal questionnaire in 1989, a telephone interview in 1993–1994, and provided a blood sample in 1993–1996 (Whitfield et al., 1998). Nearly all are of European descent, with the majority having British or Irish ancestry. Although all were twins, in some cases only one member of a twin pair provided blood. We determined zygosity from responses to questions about physical similarity and the inability of others to tell them apart, supplemented by blood group information and, for those DZ pairs participating in linkage projects (Beekman et al., 2003), genome-wide microsatellite genotyping. Participants gave written informed consent and the studies were approved by the appropriate ethics committees.

Blood was collected from 1,134 men and 2,241 women. At the same visit, their height and weight were measured and body mass index (BMI) was calculated as weight (kg)/[height (m)]^2. Information on alcohol intake (the number of drinks in the previous week) was obtained by self-report questionnaires. Information on smoking was derived from the 1989 questionnaire, and its use for the 1993–1995 period has been validated in a previous paper (Whitfield et al., 2000). Data on the number of years of education (in seven categories) and social class (in three categories) were extracted from self-reports in the 1989 questionnaire (Miller et al., 2001). Participants' addresses were categorized (using their postcodes at the time of blood collection and a database on the geographic location of Australian postcodes) into urban, suburban, or rural zones. Information on the subjects for whom results are available is summarized in Supplementary Table 1.

Laboratory procedures. Blood was collected into EDTA, lithium heparin, and plain tubes. Plasma, and then the buffy coat, were removed from the anticoagulated tubes after centrifugation and the remaining red cells were stored at –20°C or below until analyzed.

Erythrocytes rather than whole blood were used for elemental analysis because the original samples had been separated to maximize the amounts of plasma and buffy coat to be used for other purposes.

Before analysis, the erythrocytes were thawed at room temperature and diluted 1:20 in ammonia/ EDTA solution containing rhodium as an internal standard. As, Cd, Cu, Hg, Pb, Se and Zn concentrations were measured by inductively coupled plasma mass spectrometry (ICP-MS) on a Perkin-Elmer Elan 5000 (PerkinElmer Inc, Wellesley, MA, USA) or Varian UltraMass (Varian Inc, Palo Alto, CA, USA). Hemoglobin concentration was then measured on the diluted samples using the cyanmethemoglobin method.

Quality control (QC) samples at two levels were run at least once on each day. Analytical variation was assessed by calculating the coefficients of variation ((standard deviation/mean)x100) for the QC materials used, and by repeatability of the participants' results for 122 paired erythrocyte samples which had been obtained at the same venepuncture but processed and stored in separate tubes and measured in different batches. The median coefficients of variation were for As 14.8%, Cd 74%, Cu 7.2%, Hg 32%, Pb 14.2%, Se 14.1% and Zn 10.3%; the correlation coefficients for the duplicated analysis of participants' samples were As 0.88, Cd 0.85, Cu 0.65, Hg 0.71, Pb 0.96, Se 0.53, and Zn 0.84.

Data analysis. A total of 2,926 individuals had erythrocyte element concentrations measured. Assays were carried out on 102 different days. To adjust for daily assay variation, a data group was added to the analysis in Mx (Neale, 1999) estimating day effects from the results of the QC samples. These were equated back to daily deviations specified in the means model of monozygotic (MZ) and dizygotic (DZ) twins. This allows the error variation inherent in measurement of element concentrations in the samples to be taken into account, with both QC and twin samples contributing to the maximum-likelihood estimation of and adjustment for day effects.

Model fitting to test for effects of genetic and environmental sources of variation and multivariate analysis of covariate effects were performed using Mx. This allows simultaneous modeling of both fixed effects on the mean (e.g. from age, sex, and smoking status) and random effects on variances and covariances (e.g., additive genetic, shared, and nonshared environmental sources of variance). Use of Mx also overcomes potential statistical problems that can arise with analysis of data from related subjects, in this case twins. The Mx data analysis was performed on results from 2,832 people. There were 428 MZ female, 165 MZ

male, 218 DZ female, 90 DZ male, and 222 DZ opposite-sex pairs, and 356 female and 230 male individuals whose co-twin did not participate. The latter contribute to estimates of means and variances, but not directly to co-twin correlations.

Effects of covariates. Adjustments were made for sample hemoglobin concentration, effects of day-to-day analytical variation (incorporating QC data and batch effects assessed from the mean for all samples analysed within a batch), sex, and age. At each step, the improvement in goodness-of-fit was assessed by the likelihood ratio chi-square test as each of the variables was added. As anticipated, significant effects were found for day-to-day analytical variation, sample hemoglobin concentration, age, and sex (all $p < 10^{-3}$). We next tested for the effects of personal, geographic, and social characteristics expected to affect the results. These were sex, age, alcohol intake, smoking status, residential location (urban, suburban or rural), social class (in three categories), educational level (seven categories), and BMI. Univariate analysis using SPSS was supplemented with Mx analysis to assess the independent effects of these covariates and to take account of twin relatedness. In this, the initial or baseline model contained all the covariates; a series of submodels were then fitted in which one of the covariates was removed and the change in goodness-of-fit calculated to determine the significance of their independent contribution.

Genetic and environmental sources of variation. Since the simultaneous adjustment for all the covariate effects was computationally intensive, we saved residuals from the model containing all the covariates. These residuals were used to fit models of genetic and environmental sources of variation and covariation to the within- and between-pair covariances by zygosity group, as previously described (Whitfield et al., 2007), and also for the linkage analysis. Although in some cases MZ-pair correlations were more than twice the DZ-pair correlations, consistent with some non-additive genetic effects, these were not significant and for consistency (and because we particularly wish to distinguish between shared environmental and genetic effects) we initially tested models with additive genetic (A), nonshared environmental (E), and shared environmental (C) sources of variation. These ACE models were then compared with models containing only A and E, to determine whether shared environmental sources of variation could be excluded.

Linkage Analysis. We performed linkage scans for erythrocyte element concentrations on 501 DZ twin pairs with phenotypic data and linkage marker data. DNA was extracted from blood or buccal swabs according to standard procedures (Miller et al., 1988). Genotyping data were collated from genome-scans which had previously been done for other projects by

the Mammalian Genotyping Service, Marshfield WI; Leiden University Medical Centre, Netherlands (Beekman et al., 2003); Sequana Inc.; Gemini plc.; and the Australian Genome Research Facility. Familial relationships were verified using GRR (Abecasis et al., 2001). After errors were resolved, Mendelian segregation inconsistencies were identified and removed with SIBPAIR 0.99.9. Genotypes associated with unlikely recombination events were subsequently flagged and deleted with MERLIN 0.10.1 (Abecasis et al., 2002). The marker genetic positions were interpolated via locally weighted linear regression from the NCBI Build 34.3 physical map positions and the published "Rutger's" (Kong et al., 2004) genetic map (see http://www.qimr.edu.au/davidD/#map). Genetic positions are expressed in Kosambi cM.

Of the 501 available sib-pairs, 487 (97%) were genotyped at 300-1717 markers (mean 743) and 14 (3%) at fewer than 300 markers. Sibling IBD sharing was estimated via multipoint methods on a 5 cM grid and maximum likelihood univariate variance components linkage analyses performed as implemented in MERLIN, using residuals after adjustment for all the tested covariates. At every 5 cM, a 1-df LOD score was computed which is distributed as a 50:50 mixture of a point probability mass at 0 and a χ_1^2, being equivalent to the original parametric LOD score proposed by Morton (Morton, 1955).

Empirical Genome-wide Thresholds. For the Pb linkage peak, the probability of the observed LOD score occurring by chance was estimated through use of 1000 gene-dropping simulations as described by Abecasis et al. (Abecasis et al., 2004). For each simulation, we used MERLIN to generate a new dataset with the original phenotype but with new genotypes simulated under the null hypothesis of no linkage for all autosomal markers, retaining the same allele frequencies, marker spacing and missing data pattern. Multipoint IBDs for each simulation replicate were then computed and linkage analysis performed as described above. We recorded the highest peak observed for each chromosome and counted the number of chromosomes exhibiting a LOD score equal to or greater than a given threshold p. The empirical genome-wide probability of the observed linkage was estimated as the number of times LOD scores above the observed value occurred, divided by the 1000 genome-wide simulations.

REFERENCES TO SUPPLEMENTAL MATERIAL

Abecasis GR, Burt RA, Hall D, Bochum S, Doheny KF, Lundy SL, Torrington M, Roos JL, Gogos JA, Karayiorgou M (2004) Genomewide scan in families with schizophrenia from the founder population of Afrikaners reveals evidence for linkage and uniparental disomy on chromosome 1. Am.J.Hum.Genet. 74(3):403-417.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2001) GRR: graphical representation of relationship errors. Bioinformatics. 17(8):742-743.

Abecasis GR, Cherny SS, Cookson WO, Cardon LR (2002) Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat.Genet. 30(1):97-101.

Beekman M, Heijmans BT, Martin NG, Whitfield JB, Pedersen NL, DeFaire U, Snieder H, Lakenberg N, Suchiman HE, de Knijff P, Frants RR, van Ommen GJ, Kluft C, Vogler GP, Boomsma DI, Slagboom PE (2003) Evidence for a QTL on chromosome 19 influencing LDL cholesterol levels in the general population. Eur.J.Hum.Genet. 11(11):845-850.

Kong X, Murphy K, Raj T, He C, White PS, Matise TC (2004) A combined linkage-physical map of the human genome. Am.J.Hum.Genet. 75(6):1143-1148.

Miller P, Mulvey C, Martin NG (2001) Genetic and environmental contributions to educational attainment in Australia. Economics of Education Review 20:211-224.

Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16(3):1215.

Morton NE (1955) Sequential tests for the detection of linkage. Am J Hum Genet 7(3):277-318.

Neale MC (1999) Mx: Statistical Modeling, 5th edition. .Richmond, VA, Department of Psychiatry, MCV.

Whitfield JB, Dy V, McQuilty R, Zhu G, Montgomery GW, Ferreira MA, Duffy DL, Neale MC, Heijmans BT, Heath AC, Martin NG (2007) Evidence of genetic effects on blood lead concentration. Environ Health Perspect 115(8):1224-30.

Whitfield JB, Fletcher LM, Murphy TL, Powell LW, Halliday J, Heath AC, Martin NG (1998) Smoking, obesity, and hypertension alter the dose-response curve and test sensitivity of carbohydrate-deficient transferrin as a marker of alcohol intake. Clin.Chem. 44(12):2480-2489.

Whitfield JB, Pang D, Bucholz KK, Madden PA, Heath AC, Statham DJ, Martin NG (2000) Monoamine oxidase: associations with alcohol dependence, smoking and other measures of psychopathology. Psychol.Med. 30(2):443-454.

Supplemental Material, Table 1. Descriptive information on participants.

		Numbers of subjects and percentages of those providing data, or mean ± SD and range	
		Male	Female
N		1001	1925
Age (years)		45.0 ± 11.1 (30 to 86)	46.5 ± 12.0 (30 to 92)
Drinks in previous week	None	219 (24%)	761 (42%)
	1–7	303 (33%)	729 (40%)
	8 –14	173 (19%)	214 (12%)
	15–21	126 (14%)	69 (4%)
	22–28	35 (4%)	26 (1.4%)
	> 28	63 (7%)	9 (0.5%)
Smoking	Smokers	190 (20%)	376 (20%)
	Nonsmokers	748 (80%)	1469 (80%)
Residence	Urban	125 (13%)	198 (11%)
	Suburban	441 (47%)	772 (43%)
	Rural	374 (40%)	821 (46%)
Education	< 7 years	9 (1%)	22 (1%)
	8–10 years	146 (16%)	581 (31%)
	11–12 years	168 (18%)	463 (25%)
	Apprenticeship, diploma	209 (22%)	279 (15%)
	Technical/teachers' college	126 (13%)	262 (14%)
	University, first degree	175 (19%)	142 (8%)
	University, postgraduate training	106 (11%)	96 (5%)
Social class	Working	315 (32%)	612 (32%)
	Middle	508 (52%)	1011 (53%)
	Upper	157 (16%)	285 (15%)
BMI		25.8 ± 3.5 (17.5 to 42.6)	25.0 ± 4.7 (13.9 to 49.5)

Supplemental Material, Table 2. Covariate effects on element concentrations. Significant effects are shown as t-statistics and significance level, others as NS (not significant, $p > 0.05$).

	Cu	**Zn**	**Se**	**As**	**Pb**	**Hg**	**Cd**
Sex	-3.69***[a]	NS	-3.41**[a]	NS	10.45***[b]	NS	-5.12***[a]
Age	NS	NS	2.06*[c]	NS	4.23***[c]	NS	NS
Smoking	NS	NS	-5.73***[d]	NS	6.31***[e]	-2.86**[d]	30.28***[e]
Alcohol intake	NS	NS	4.50***[f]	4.35***[f]	8.70***[f]	4.83***[f]	NS
BMI	NS	NS	NS	NS	NS	NS	NS
Social class	NS	NS	NS	NS	NS	NS	NS
Education	NS	NS	NS	-2.05*[g]	-3.00**[g]	NS	-2.25*[g]
Residence	NS	-2.22*[h]	NS	NS	NS	3.61***i	NS

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

[a] Higher concentrations in women, [b] Higher concentrations in men.
[c] Higher concentrations in older people.
[d] Higher concentrations in non-smokers, [e] higher concentrations in smokers.
[f] Higher concentrations in those who drink more alcohol.
[g] Higher concentrations in those with less education.
[h] Higher concentrations in those with non-urban residence, [i] higher concentrations in those with urban residence.

Supplemental Material, Table 3. Results of multivariate analysis of genetic and environmental variation and covariation for the seven elements. Within each triangular matrix, values on the diagonal show the proportions of variance unique to a single element while values below the diagonal show the proportions of covariance shared by the two elements for that column and row.

Additive genetic effects (A)

	Cu	**Zn**	**Se**	**As**	**Pb**	**Hg**	**Cd**
Cu	26						
Zn	6	20					
Se	8	2	12				
As	0	0	0	15			
Pb	2	0	0	0	35		
Hg	0	0	1	18	0	7	
Cd	0	0	0	0	3	0	14

Shared environmental effects (C)

	Cu	**Zn**	**Se**	**As**	**Pb**	**Hg**	**Cd**
Cu	1						
Zn	3	11					
Se	0	1	9				
As	2	0	1	0			
Pb	2	0	0	0	0		
Hg	0	0	0	1	0	1	
Cd	0	0	0	0	0	0	6

Non-shared environmental effects (E)

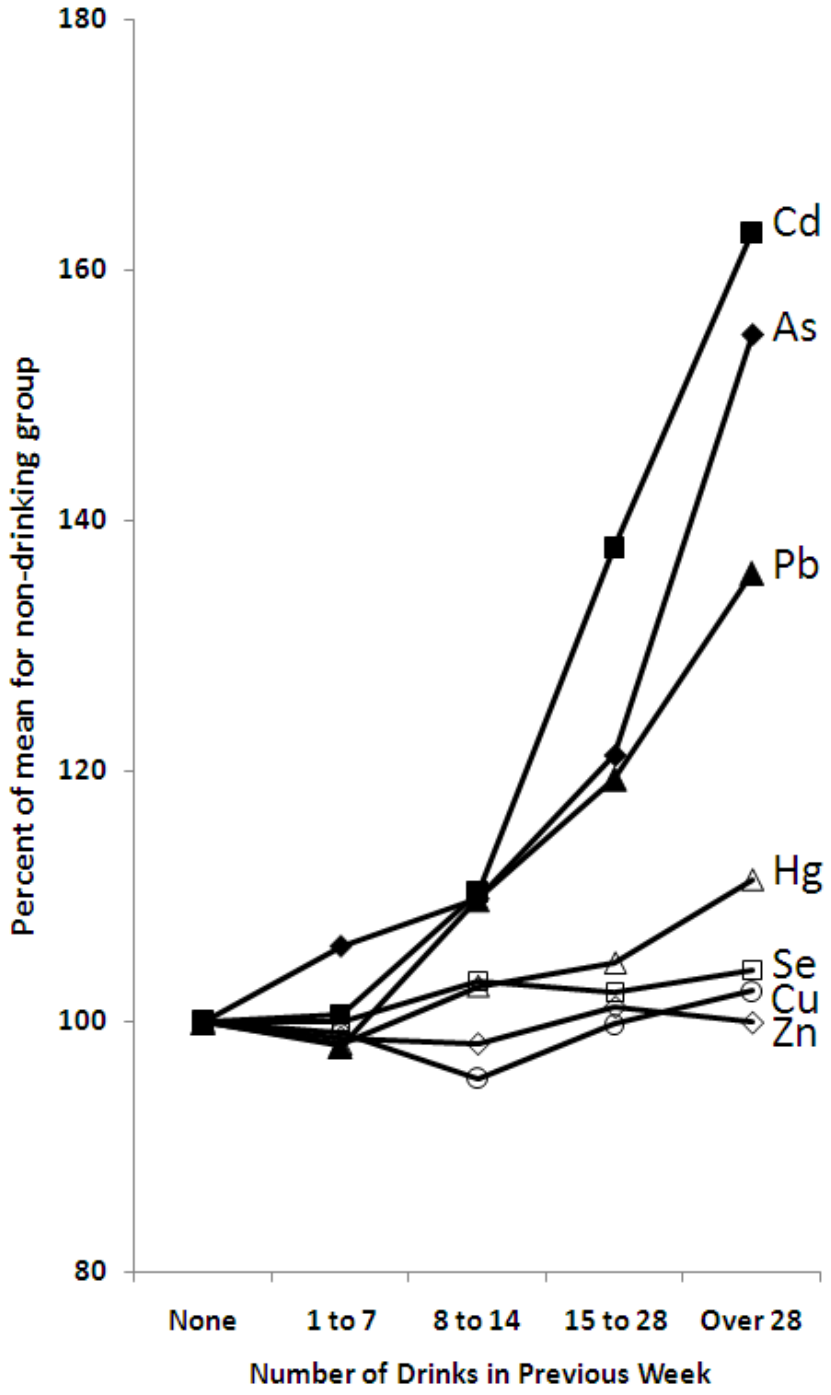
	Cu	**Zn**	**Se**	**As**	**Pb**	**Hg**	**Cd**
Cu	72						
Zn	32	29					
Se	18	7	43				
As	2	0	4	76			
Pb	4	1	0	0	56		
Hg	0	0	6	5	0	59	
Cd	0	0	0	0	0	0	76

Supplemental material, Table 4. Summary of linkage results, showing all loci where individual elements have LOD ≥ 1.6, and immediately surrounding regions. Physical distances (Mb) are estimated using the integrated genetic map at http://www.qimr.edu.au/davidD/#map .

Chr	cM	Mb	Cu	Zn	Se	As	Pb	Hg	Cd
2	120	106	0.14	0.52	0.00	0.02	0.04	0.00	**1.88**
2	125	113	0.02	0.48	0.00	0.01	0.03	0.00	**1.75**
2	210	212	0.42	**1.78**	0.43	0.00	0.06	0.00	1.14
2	215	215	0.27	**2.41**	0.82	0.00	0.10	0.00	1.26
2	220	219	0.36	**2.73**	1.25	0.00	0.05	0.00	1.27
2	240	232	0.00	0.89	0.20	0.00	0.00	0.00	**1.82**
2	245	233	0.00	0.56	0.03	0.00	0.00	0.00	**2.04**
2	250	237	0.00	0.33	0.00	0.00	0.04	0.00	**1.98**
2	255	239	0.00	0.29	0.00	0.00	0.08	0.00	**1.97**
2	260	241	0.00	0.17	0.00	0.00	0.03	0.00	**1.84**
2	265	242	0.00	0.15	0.00	0.00	0.01	0.00	**1.64**
3	40	18	0.00	0.00	0.09	0.30	**1.81**	0.00	0.00
3	45	23	0.00	0.00	0.07	0.27	**2.97**	0.00	0.00
3	50	27	0.00	0.00	0.03	0.27	**4.21**	0.00	0.00
3	55	31	0.00	0.00	0.00	0.26	**4.08**	0.00	0.00
3	60	35	0.00	0.18	0.00	0.12	**3.77**	0.00	0.00
3	65	41	0.00	0.01	0.01	0.08	**2.68**	0.00	0.00
3	70	49	0.00	0.00	0.13	0.00	**1.87**	0.00	0.00
3	75	56	0.00	0.00	0.07	0.00	**1.69**	0.00	0.05
3	80	59	0.02	0.00	0.07	0.00	**1.88**	0.00	0.01
4	125	120	1.14	0.13	**1.92**	0.00	0.05	0.00	0.15
4	130	128	1.28	0.48	**2.46**	0.00	0.03	0.00	0.43
4	135	135	0.79	0.35	**1.70**	0.19	0.10	0.00	0.56
5	160	152	0.17	0.31	0.64	0.68	0.14	1.35	0.31
5	165	158	0.30	0.30	0.56	0.68	0.46	1.28	0.98
5	170	163	0.47	0.26	0.39	0.69	0.44	1.40	0.75
5	175	166	0.47	0.36	0.30	0.64	0.42	**1.65**	0.46
5	180	169	0.55	0.47	0.10	0.45	0.42	1.38	0.24
8	15	6	0.09	0.71	1.59	0.00	0.70	0.07	0.00
8	20	9	0.20	0.75	**1.76**	0.00	0.70	0.16	0.00
8	25	11	0.38	1.09	**2.05**	0.04	0.53	0.40	0.00
8	30	15	0.61	0.95	**1.88**	0.34	0.92	0.08	0.00
8	35	18	0.88	1.01	**1.84**	0.46	1.11	0.01	0.00
8	40	20	0.99	0.92	1.51	0.56	1.16	0.00	0.00
8	45	23	0.77	0.71	1.32	0.70	0.64	0.00	0.00
8	50	26	0.67	0.60	1.21	0.69	0.24	0.00	0.00

18	0	1	0.04	0.14	0.31	0.07	0.00	0.00	**2.00**
18	5	2	0.04	0.26	0.29	0.09	0.00	0.00	**1.81**
18	20	6	0.09	0.03	0.15	0.00	0.00	0.00	**1.72**
18	25	8	0.36	0.08	0.30	0.00	0.00	0.00	**1.81**
20	10	3	0.00	0.00	0.00	0.00	0.00	0.00	**1.80**
20	15	5	0.00	0.00	0.01	0.00	0.00	0.00	**2.02**
20	20	6	0.00	0.00	0.09	0.00	0.00	0.00	**1.86**
20	25	8	0.00	0.00	0.13	0.00	0.00	0.00	**1.89**
20	30	10	0.00	0.00	0.03	0.00	0.00	0.00	**2.05**
20	35	12	0.00	0.00	0.00	0.00	0.12	0.00	**1.74**
X	95	91	0.00	0.00	0.00	0.00	0.00	0.00	**1.79**
X	100	97	0.00	0.00	0.00	0.00	0.00	0.00	**1.63**

Supplemental Material, Figure 1. Effects of self-reported alcohol intake on erythrocyte element concentrations. Results are adjusted for analytical variables, age and sex but not smoking status. For each element, the means for each drinking group have been adjusted to percentage of the mean for those reporting no drinks in the previous week.



Supplemental Material, Figure 2. Results of genetic linkage analysis for each of the seven elements. X-axis: chromosomes and distance along each chromosome. Y-axes: LOD scores.